

The Cinderella discipline: morphometrics and their use in botanical classification

Article

Accepted Version

Christodoulou, M. D. ORCID: <https://orcid.org/0000-0003-0968-2311>, Clark, J. Y. and Culham, A. ORCID: <https://orcid.org/0000-0002-7440-0133> (2020) The Cinderella discipline: morphometrics and their use in botanical classification. *Botanical Journal of the Linnean Society*, 194 (4). pp. 385-396. ISSN 0024-4074 doi: 10.1093/botlinnean/boaa055 Available at <https://centaur.reading.ac.uk/91282/>

It is advisable to refer to the publisher's version if you intend to cite from the work. See [Guidance on citing](#).

To link to this article DOI: <http://dx.doi.org/10.1093/botlinnean/boaa055>

Publisher: Oxford University Press

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the [End User Agreement](#).

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

1 **REVIEW**

2 TITLE: The Cinderella Discipline: Morphometrics and their use in botanical classification

3

4 SHORT TITLE: Morphometrics and their use

5

6 AUTHORS: Christodoulou, M.D.^{1,2*}, Clark, J.Y.³, and Culham, A.²

7

8 ¹ Department of Statistics, University of Oxford, Oxford, UK

9

10 ² University of Reading Herbarium, School of Biological Sciences, University of Reading,

11 Whiteknights, Reading, UK

12

13 ³ Department of Computer Science, Faculty of Engineering and Physical Sciences,

14 University of Surrey, Guildford, UK

15

16 * For correspondence e-mail: maria.christodoulou@stats.ox.ac.uk

17

18 **ABSTRACT**

19

20 Between the 1960s and the present day, the use of morphology in plant taxonomy
21 suffered a major decline, in part driven by the apparent superiority of DNA-based
22 approaches to data generation. However, in recent years computer image
23 recognition has re-kindled the interest in morphological techniques. Linear or geometric
24 morphometric approaches have been employed to distinguish and classify a wide
25 variety of organisms; each has strengths and weaknesses. Here we review these
26 approaches with a focus on plant classification and present a case for the
27 combination of morphometrics with statistical/machine learning. There is a large
28 collection of classification techniques available for biological analysis and selecting the
29 most appropriate one is not trivial. Performance should be evaluated using
30 standardised metrics such as accuracy, sensitivity, and specificity. The gathering and
31 storage of high-resolution images, combined with the processing power of desktop
32 computers, makes morphometric approaches practical as a time- and cost-efficient
33 way of non-destructive identification of plant samples.

34

35 Keywords: Plant taxonomy, geometric morphometrics, linear morphometrics, statistical
36 learning, machine learning, identification, classification, neural networks.

37

38 In his keynote address during the 50th anniversary of botany MSc training at the
39 University of Reading, Prof Vernon Heywood described a steady decline in the state of
40 botany teaching in the UK with a resulting loss of skills in the next generation of scientists.
41 With few institutions in the country offering training for young botanists, more and more
42 researchers enter plant taxonomy through the field of molecular systematics, never
43 learning the classic skills of a traditional botanist. Although great progress has been
44 made in the development of molecular tools, increasing the insight gained from
45 laboratory methods, what used to be the beating heart of botany - morphology - has
46 lost some of its appeal. In our view this is because morphological data coding cannot
47 readily be made into a clear data generation pipeline in the same way as much
48 molecular data can. We believe this to be because morphology requires more in-
49 depth knowledge and understanding of the organism prior to data collection than is
50 required for DNA sequencing and that morphological variation is open-ended rather
51 than with a fixed range of states as in DNA data. Whilst morphological data have lost
52 favour in the construction of plant classification systems they have gained popularity in
53 the study of evolution from variation in gross morphology of the centropogonid clade
54 (Lobelioideae: Campanulaceae) (Lagomarsino *et al.*, 2017), speciation despite
55 consistent floral morphology in *Myrcia* DC. (Vasconcelos *et al.*, 2019) though to
56 detailed morphometric analysis of traits related to environment in *Vriesea* Lindl.
57 bromeliads (Neves *et al.*, 2020).

58

59 The power of some of the more modern developments in morphometrics and statistical
60 learning however can provide botanists with an extra toolbox to help them describe
61 and quantify the variation that surrounds them. In this review we aim to make a case
62 for the value of morphometrics, especially in combination with more sophisticated
63 statistical methods, in a botanist's analytical toolbox - not to replace molecular
64 techniques but to add to them. Morphology is often one of the most directly accessible
65 and intuitive data sources for taxonomic research. In botanical taxonomy,

66 morphological characterization is the foundation of taxon description and
67 identification, albeit often found in the formal and stylised format present in Floras and
68 monographs. There is an opportunity for modern botanical taxonomy to explore the
69 rapidly advancing field of morphometrics which already has some notable examples
70 ranging from automatic leaf outline identification of *Passiflora* L. species (De Oliveira
71 Plotze & Martinez Bruno, 2009), the tooth margin algorithm for *Tilia* L. leaf identification
72 (Corney *et al.*, 2012), and the use of leaf venation architecture for major angiosperm
73 clade recognition (Wilf *et al.*, 2016). Some computerised systems, starting with an
74 existing classification of taxa, can use machine learning to handle the routine
75 identification work, and then refer intransigent problems to a human expert (Clark,
76 Corney, & Wilkin, 2017).

77

78 One of the principal arguments presented against morphological data is the potential
79 for high levels of ambiguity. This ambiguity can be caused by a variety of factors such
80 as inaccurate character definition (Assis, 2009) and difficulty in establishing homology
81 (Schneider, Smith, & Pryer, 2009). Morphological data collection can be further
82 complicated by plasticity of features (Perkins, Martinsen, & Falk, 2011), homoplasy
83 (Schneider *et al.*, 2009), low numbers of characters (Giribet, 2010), and missing
84 character states (Jenner, 2004). In some organisms such as parasites, reduced body
85 plans can make characterisation of features even more difficult and lead to a very
86 limited dataset (Perkins *et al.*, 2011). These concerns are neither exaggerated nor trivial
87 and many are thoroughly discussed in the morphological literature. They do not,
88 however, necessarily imply a lower quality of data produced by morphological work in
89 comparison with other data sources (Jenner, 2004).

90

91 As most botanical researchers are question-driven rather than method-driven, we have
92 structured our recommendations using general outlines on what kind of questions each
93 combination of morphological and statistical tools can answer, with the aim of

94 promoting more thorough morphological investigation in botanical research. We have
95 split this into two sections - Developmental hypotheses and Classification hypotheses.
96 Under Developmental hypotheses we include all studies that may require the
97 description of shape or size of a plant either to compare between treatments or to
98 study how characters change along a particular gradient. For these we give an outline
99 of morphometric tools available. Under Classification hypotheses we include all studies
100 where the researcher is asking questions of taxon membership (e.g. are these two
101 groups in the same taxon?) or questions of identification (e.g. what is the minimum set
102 of diagnostics to accurately identify a sample?). These also require morphometric tools,
103 such as those described under the developmental hypotheses, but can be taken
104 further by combining them with machine learning techniques. There is a difference in
105 terminology between the use of the word classification in biology and in computer
106 science. Although the term is clearly defined in a taxonomic setting as the formal
107 structure in which taxa are placed, in machine learning it means something much
108 more general: it is the attribution of objects to a particular group. This is why
109 identification in the machine learning context falls under classification, and therefore is
110 included here under Classification hypotheses.

111 **DEVELOPMENTAL HYPOTHESES**

112 Plant growth and development studies already rely heavily on morphological
113 measurements - size for example is often included as a proxy to an organism's
114 developmental stage. These studies often focus on examining how the organism
115 changes as it progresses through the various life stages. These could range from
116 progression from seed to flower for an annual, or even development of fruit on a tree
117 during the growing season.

118

119 Even though it is very commonly used, size itself is a complex and often unappreciated
120 concept. Often researchers fail to explore the separation between shape and size,
121 confounding the two and losing some of the clarity that can be obtained through their
122 investigation. For Developmental hypotheses, we argue that the crucial point for insight
123 is not the separation of size and shape just for the sake of it - it is for the researcher to
124 either knowingly combine them or distinguish between the two based on the
125 hypothesis in question. We believe that by cautiously selecting measurements that do
126 not distinguish shape from size, a researcher can gain insight on changes in either size
127 or shape during a developmental process based on how they use them. For example,
128 the length of apple fruit through the growing season, plotted against time from anthesis
129 can give insight on how size develops as time progresses (Atay, Pirlak, & Atay, 2010).
130 The ratio between length and width for the same fruit provides an indication of the
131 development of shape (Bollard, 1970). Both length and width independently are size
132 metrics, but in combination they describe shape.

133

134 In the context of describing morphology, there are two mainstream methods: linear
135 and geometric morphometrics. An essential distinction between them is that linear
136 morphometrics do not actively separate size from shape, whereas geometric
137 morphometrics do. We have structured the remainder of this section to describe these
138 two techniques and have illustrated them using biological examples.

139

140 The traditional approach to morphometrics involves the measurement of distances
141 between points deemed to be characteristic of shape and form. Measurements such
142 as height, length, width, and diameter all fall under the general categorisation of **linear**
143 **morphometrics**. These measurements are intuitive, easy to understand and to interpret,
144 and have been in the biological toolbox for as long as the toolbox itself has existed.
145 Linear morphometrics are quick to collect, low cost, easy to interpret, and often
146 sufficient for biological description. Sanchez et al. (2011) compared the growth
147 development of baobab seedlings of different origins using a variety of morphometric
148 measurements ,such as length and diameter of roots, to establish that plants originating
149 from drier environments grew to a smaller size even under optimal greenhouse
150 conditions. Richardson et al. (2011) the studied fruit development patterns of kiwi from
151 anthesis to ripening using amongst other character a collection of linear
152 morphometrics, such as pericarp diameter. Zhang et al. (2015) performed a
153 comparative study of the developmental patterns of Sweet cherry floral parts, using
154 linear measurements such as pedicel length, establishing a correlation between floral
155 morphology and environmental conditions during growth such as temperature.

156

157 For morphometric studies that require the description of very subtle shape characters,
158 linear morphometrics may not be the most appropriate tool. The reason for this is
159 because distance measurements, although excellent for summarising shape and size
160 descriptions, often lack context. To correct for this, more linear measurements can be
161 collected, creating a more complete dataset for each object. When the shape of
162 interest is of biological form, it becomes crucial to be able to establish and quantify
163 even the subtlest of differences. To be able to achieve this through linear
164 morphometrics would involve an extensive collection of measurements and a
165 generous amount of luck, as one may simply fail to measure the precise point where
166 differences between taxa occur. Furthermore, within an evolutionary framework it is

167 more appropriate to view form as a whole since organisms evolve as a whole. To
168 counter these concerns, morphometric theory progressed to what is often described as
169 modern morphometrics, more accurately known as geometric morphometrics.

170

171 **Geometric morphometrics** allow the study of the shape of an organism as a whole,
172 rather than as a collection of separate components. In contrast to linear
173 morphometrics, by studying all the selected landmarks of a sample together, even
174 subtle changes in geometry can be quantified and analysed using geometric
175 morphometrics. Kendall's shape definition forms the basis of geometric morphometrics
176 (Zelditch *et al.*, 2004). This clear separation of shape from position, orientation and size
177 corresponds to an intuitive concept of shape. In practical terms, to achieve this
178 separation there is a strong analytical reliance on multivariate techniques (Klingenberg
179 & Monteiro, 2005). The way this is performed in geometric morphometrics is through the
180 use of landmark coordinates (Van Bocxlaer & Schultheiß, 2010). A landmark is a
181 recognizable point on the organism that, together with other landmarks, can be used
182 to summarise the form of the organism (Zelditch *et al.*, 2004). As opposed to focusing
183 on distance measurements, as is done in linear morphometrics, shape is summarised
184 through the Cartesian coordinates of selected landmarks (Walker, 2000). By always
185 analysing these coordinates together in a multidimensional space, shapes can be
186 scaled, moved and rotated without losing any information (Goodall, 1991). Although
187 the selection of appropriate landmarks can be difficult, this multivariate approach
188 provides great flexibility for manipulation and statistical analysis.

189

190 After landmark selection, the recording of coordinates for all samples (a process
191 referred to as "sample digitisation") creates the initial dataset to be used for analysis.
192 The samples in this dataset are not, however, comparable if their coordinates have not
193 been standardised. This is because regardless of how carefully and methodically
194 digitisation occurred the samples are bound to not be fully aligned. Furthermore,

195 differences in sizes between samples will affect the position of the landmarks on the
196 Cartesian axes, confounding shape comparison. To correct for this, the samples can be
197 standardised using a Procrustes superimposition (Rohlf & Slice, 1990). Named after the
198 mythical ancient Greek bandit who trimmed or stretched his victims to fit an iron bed,
199 the process superimposes the samples using the landmarks to correct for orientation
200 and alignment (Stegmann & Gomez, 2002). It then proceeds to stretch or shrink some
201 samples aiming for all samples to be perfectly superimposed (Zelditch *et al.*, 2004). We
202 have illustrated the steps of this process in Figure 1.

203

204 Selecting appropriate landmarks to summarise a shape is perhaps the most crucial
205 aspect of geometric morphometrics. The reason for this is that if the choice of
206 landmarks is poor, then any subsequent analysis will reflect that. Through the process of
207 landmark selection, the overall shape of the organism in question is summarised using a
208 small number of representative landmarks. Selecting representative landmarks is a
209 subjective exercise that relies on in-depth knowledge and understanding of anatomy
210 and biology of the organism in question. This is because not all landmarks are created
211 equal. A wisely chosen landmark can summarise shape appropriately and provide
212 adequate information for biological inference. A poorly selected landmark will at best
213 add high levels of noise to the dataset or, at worst, result in misleading patterns.

214

215 Ideally, landmark selection requires four criteria that ensure quality: **repeatability**,
216 **consistency of position**, **adequacy** and **homology** (Zelditch *et al.*, 2004). **Repeatability**
217 refers to the potential of locating the selected landmark accurately on a specimen
218 multiple times (Zelditch *et al.*, 2004). If a landmark is difficult to locate or its position is
219 relatively vague, then samples that have no significant biological differences may be
220 found to be different as an artefact of poor landmark choice. **Consistency of position**
221 refers to the relative positions between landmarks (Zelditch *et al.*, 2004). If two
222 landmarks switch relative positions between different specimens then their comparison

223 can lead to statistical outliers that may affect the findings and analyses. **Adequacy**
224 refers to the number and position of landmarks used to summarise a form (Zelditch *et*
225 *al.*, 2004). Although more is not always better in terms of landmark selection, including
226 too few landmarks will not lead to a representative dataset. Even though repeatability
227 can be quantified, and consistency of position detected, adequacy is a harder
228 criterion to evaluate. This is because adequate coverage can be highly subjective. The
229 concept relies on finding the golden mean between oversampling the specimen
230 (where too many landmarks can lead to higher noise levels in the dataset) and
231 undersampling (losing possible detectable variation between specimens).

232

233 **Homology** in landmark selection has both geometric and biological aspects. Two
234 landmarks are considered homologous in two specimens if there is a degree of
235 correspondence between them. This correspondence can be purely a geometric
236 attribute (e.g. the tips of the Giza pyramids are geometrically homologous) or a
237 biological attribute (e.g. the forelimbs of bats and primates). Although all four criteria
238 are important for landmark quality, establishing homology is crucial. It is only through
239 the use of homologous landmarks that the shapes studied are truly comparable. If the
240 landmarks used are not homologous between the organisms in the study then there is
241 no logical support for their comparison and the results can be highly misleading
242 (Klingenberg, 2008). Although homology is considered one of the most crucial aspects
243 in landmark selection, exactly how it can affect a given study depends on the nature
244 and scope of the study itself. In general, the ability to identify homology can severely
245 limit the quantity of potential landmark candidates.

246

247 These constraints imposed by homology increase the popularity of outline methods of
248 analysis (Macleod, 1999). By replacing homologous landmarks with regularly spaced
249 points along a curve, outline analysis sidesteps the issue of homology and can be used
250 in cases where landmarks are sparse or hard to define (Macleod, 1999). Outline data

251 can then be analysed using Fourier harmonics (or possible variations such as Elliptical
252 Fourier) or Eigenshape analysis (Macleod, 1999; Bonhomme & Claude, 2014). Although
253 outline analysis is a popular and successful alternative to landmark analysis, the
254 assumption that it bypasses homology issues may be misplaced. The reason for this is
255 that outline methods are not completely independent of landmark correspondence
256 assumptions (Klingenberg, 2008). That is because as with landmark methods, outline
257 coordinates require a superimposition technique, such as Procrustes superimposition,
258 prior to analysis (Bonhomme & Claude, 2014). This means that the outline points that
259 are recorded are treated as actual homologous landmarks. This may appear minor,
260 but as the superimposition process assumes a certain correspondence between points
261 on the outline, it can result in increased levels of noise in the dataset. Furthermore,
262 analytical approaches such as Elliptic Fourier Analysis also assume a certain degree of
263 homology between outline points. It can therefore be argued that the principal
264 difference between the two approaches is that in landmark analysis the homology
265 criterion is explicit whereas in outline analysis it is implied and often ignored.

266

267 The choice between linear and geometric morphometrics for an analysis is not trivial as
268 one technique is not necessarily superior to the other. Linear morphometrics are quick,
269 intuitive and cost effective and often robust enough to not introduce noise in the
270 analysis. They fail when separation of shape and size becomes important and when
271 subtle changes in morphology are crucial - this is where geometric morphometrics
272 excel. Selecting the appropriate method for the question in hand is always a
273 challenging aspect of scientific discovery, although familiarity with both methods,
274 combined with understanding of the studied organism helps when deciding which
275 technique may provide more insightful findings. As a final point, it is not always
276 necessary to choose one over the other, for example, Christodoulou et al. (2018)
277 combined linear and geometric morphometrics to describe shape differences
278 between apple cultivars with greater accuracy.

279 **CLASSIFICATION HYPOTHESES:**

280 Although classification in biology has a different meaning than in machine learning (a
281 subset of statistical learning), this collection of hypotheses relies on grouping objects
282 based on similarities between measured characters. These can include studies of
283 morphological similarities between geographically distinct populations, segregation
284 between species and hybrids, or revision of taxonomic limits.

285

286 Both linear and geometric morphometrics have been used for such studies. Compton
287 and Hedderson (1997), in their taxonomic revision of the limits of *Cimicifuga foetida* L.
288 s.l. (now *Actaea cimicifuga* L.), included 17 length variables, resulting in the detection
289 of four geographically distinct species. Blanco-Dios (2007) used multivariate analysis of
290 17 linear morphometric characters to contrast the morphology of hybrid populations
291 between *Armeria beirana* Franco and *A. pubigera* (Desf.) Boiss. with that of their
292 progenitors, detecting clear differences between the groupings. Da Costa et al. (2009)
293 used distance measurements for both vegetative and reproductive parts to study the
294 variation within the *Vriesea paraibica* Wawra complex. After statistical analysis, they
295 proceeded to recognise four species within the complex (*V. paraibica*,
296 *V. interrogatoria* L.B.Smith, *V. eltoniana* E.Pereira & Ivo, and *V. flava* A.F. Costa, H.
297 Luther & M.G.L. Wanderley), for which they provided a taxonomic treatment. Returning
298 to the genus *Actaea* L., Gardner et al. (2012) used linear morphometrics to quantify the
299 variation within *Actaea racemosa* L., establishing that between-population variation
300 was similar to within-population variation. In a study of the *Andropogon lateralis* Nees
301 complex, Nagahama et al. (2014) used 19 linear morphometric measurements to
302 successfully distinguish both species and hybrids within the complex. Shipunov and
303 Bateman (2005) used geometric morphometrics to explore the diversity of lip shapes of
304 *Dactylorhiza* Neck. ex Nevski orchids, studying both hybridization patterns and
305 taxonomy in Russian populations. Volkova and Shipunov (2007) used similar tools to
306 investigate the variation between three *Nymphaea* L. species in Russia and Siberia,

307 finding the species delimitation to be robust. Viscosi et al. (2009) successfully used
308 geometric morphometrics on oak leaves to distinguish between four species. Savriama
309 et al. (2012) presented a new methodology quantifying symmetry and asymmetry of
310 corolla shape in *Erysimum mediohispanicum* Polatschek (now *Erysimum grandiflorum*
311 subsp. *mediohispanicum* (Polatschek) Romo), establishing symmetry to be a
312 fundamental character for floral variation within the taxon. Finally, Fernández-
313 Mazuecos et al. (2013) used geometric morphometrics to study the role of flower
314 specialisation for speciation in *Linaria* Mill. subsect. *Versicolores* (Benth.) Wetst. finding
315 corolla tube differences to correlate with divergent pollination strategies. In a
316 comparison of leaf shape of *Anacardium microcarpum* Ducke with *A. occidentale* L.
317 using geometric morphometric descriptors, Vieira et al. (2014) established that
318 although the leaves do present statistically significant differences, overlap between
319 taxa and populations prevent them from being used as unique identifiers.

320

321 Analytically, methods from statistical/machine learning can offer great insight for this
322 type of hypothesis. There are two broad sections in statistical/machine learning:
323 supervised learning and unsupervised learning. We are excluding deep learning
324 methods here, as the topic is too large for an adequate description within this review
325 and the approaches are rather different. The review on the topic by Angermueller et
326 al. (2016) offers a good overview of the major issues. Furthermore, deep learning is
327 primarily aimed at processing huge amounts of multivariate data (so called 'big data'),
328 and here we are more concerned with the utilisation of relatively small datasets, often
329 with only a few data records per taxon, which is more realistic for consideration by
330 practising botanists.

331

332 **Supervised learning** focuses on using combinations of characters to circumscribe
333 known groups (classes) and then applying this knowledge to predict the class
334 membership of an unknown sample (Tarca et al., 2007). This is essentially 'identification'

in the biological sense, if the classes represent named taxa. The classic example of supervised learning is Anderson's *Iris* dataset analysed by Fisher using Linear Discriminant Analysis (LDA) (Fisher, 1936). The original dataset contained measurements from 150 flowers belonging to three *Iris* species (50 flowers each of *I. setosa* Pall. ex Link, *I. versicolor* Thunb. and *I. virginica* L.), For each flower, length and width measurements of two tepals (one inner, and one outer tepal), as well as species, were recorded. When this dataset was analysed using LDA, discriminant functions were established for each species based on the lengths and widths of the tepals. These could then be used to establish the species of an unknown *Iris* sample using only length and width tepal measurements (provided it belonged to one of the three species). The factor that makes this example part of supervised learning is the prior knowledge of class membership, in this case *Iris* species, used for the design of the discriminant functions (Fogel, 2008).

Unsupervised learning, by contrast, has no prior knowledge of class membership, and the analysis aims to explore patterns in the data and create natural groupings (Fogel, 2008). Such groupings can then be used as justification for delimitation of traditional ranked taxa such as species. This is essentially 'classification' in the biological sense. Cluster analysis (clustering), for example, is a case of unsupervised learning. Table 1 summarises a selection of both supervised and unsupervised techniques, more extensive descriptions of which can be found in Appendix A.

Table 1 showcases botanical applications of machine learning. The combination of machine learning and morphometrics for classification has much more prominent examples outside of botany. We aim for this review to increase the uptake of these techniques in botany. In the meantime, we present some non-botanical examples here for illustration purposes. Santana et al. (2014) studied bee classification using the forewings of male members of five *Euglossa* species. This was performed by using 18

landmarks on the wing venation together with colour change variables, followed by comparisons between classification techniques including linear discriminant analysis and a modified neural network. The neural network outperformed the other classifiers, with an accuracy of 87.6%. da Silva et al. (2015) used more classes than Santana et al. (2014), studying 26 subspecies of *Apis mellifera* while still using the same 18 landmarks on wing venation. Their focus was on the performance of feature selection and their conclusion was that a Naïve Bayes classifier outperforms other classification techniques, with 65% mean accuracy on cross-validation (da Silva et al., 2015).

Van Bocxlaer and Schultheiß's (2010) gastropod study was one of the first in zoology to combine machine learning with morphometrics, their focus was primarily on comparing landmark analysis with outline analysis. For their gastropod dataset they found that outline analysis outperformed landmark analysis by 3%, reaching 78% accuracy when using a Support Vector Machine (SVM) classification (Van Bocxlaer & Schultheiß, 2010). The high success rate of the outline analysis is likely due to the presence of three-dimensional ornamentation on the shell surface. Also, the theory of outline methods for biological shape analysis is not as robust as landmark analysis, as discussed briefly in earlier sections.

Guisande et al. (2010) describe new software designed to identify fish species, using Classification and Regression Trees (CARTs) and linear morphometrics. The structure of the software is such that the user is required to make linear measurements on their sample, following a certain protocol, and the measurements are then used to classify the sample. This makes it similar to a multi-access key rather than a tool for automatic identification. For multi-access keys, success rates can be established by testing the key on the target audience and recording how successful was their navigation of the key. Guisande et al. (2010) did not perform this test and only tested accuracy using samples they had measured themselves.

391

392 In the field of anthropology, Velemínská et al. (2013) used semi-landmarks to study the
393 greater sciatic notch (which is part of the pelvis bones) aiming to correctly classify the
394 sex of the individual. Their best performing classifier was a Support Vector Machine that
395 achieved a 92% accuracy. Instead of using a completely independent test set, the
396 accuracy was quantified using a leave-one-out cross-validation approach on the
397 learning set. The absence of a separate test set can lead to overestimating the
398 accuracy of the classification as briefly discussed earlier.

399

400 The orthodontics paper by Yu et al. (2014) is based on the unusual premise of
401 predicting attractiveness on malocclusion patients (patients with misaligned teeth). By
402 using 101 landmarks on patient images combined with a Support Vector Machine, they
403 achieved an accuracy of attractiveness prediction of 72%. This work is interesting
404 because it is the only example in the literature where geometric morphometrics have
405 been combined with the regression approaches of statistical learning, rather than the
406 classification ones. This is because the attractiveness measure used was based on a
407 (subjective) score from 69 orthodontics experts, therefore the prediction was a
408 continuous measurement rather than a class.

409

410 **Model evaluation**

411 There is a large collection of classification techniques available for biological analysis
412 and selecting the most appropriate technique is not trivial. The reason for this is that
413 there is no single classification technique that consistently outperforms all others
414 regardless of the dataset studied. In machine learning this concept is referred to as the
415 "No free lunch" Theorem. Stated formally by Wolpert and Macready (1997), the
416 theorem suggests that the performance of all classifiers is equal when the totality of
417 possible problems is considered. This means that for every classifier available there exists
418 a possible problem where that classifier outperforms every other classifier. In practical

419 terms, this makes selecting a classifier for a study harder as the only way to establish the
420 appropriateness of the technique is after the training of the classifier. Due to this, the
421 most common approach to classification problems is to train a variety of different
422 classifiers and then select the one that performs best (Fogel, 2008). This strategy makes
423 performance evaluation the focus of the classification analysis. To this extent a series of
424 metrics have been proposed in the literature, summarised in Table 2.

425

426 All the metrics presented in Table 2 rely on describing classification success through the
427 use of a set of samples, however selecting the set that is used is not straightforward. In
428 most biological situations there is a limited amount of data available for study, making
429 each individual sample valuable to the study. With a limited dataset, therefore, the
430 decision on the appropriate "spending" of the data is not an easy one to make. This
431 makes pilot studies that can inform power analyses (to estimate appropriate sample
432 sizes) a crucial aspect of experimental design (McDonald, 2014).

433

434 There are three stages in machine learning that require data: training, validating and
435 testing (Olden, Lawler, & Poff, 2008). During the first stage the classifier is primarily
436 trained to the problem in question. If the whole dataset is used at this stage then it will
437 have to be re-used for both validating and testing, leading to potential overfitting and
438 unrealistically high performance metrics (Olden *et al.*, 2008). This is because the
439 classifier would have knowledge of the full dataset at the training stage, therefore
440 when validating occurs (which is the process that verifies that appropriate tuning
441 parameters have been selected during training), overfitting is more likely as none of the
442 validating samples will be new. When the classifier is then tested using known samples,
443 the performance will appear improved due to this overfitting effect. The peril from this is
444 that when the classifier is applied to truly unknown samples, the confidence in the
445 resulting class could be misplaced. To avoid this, common practice involves partitioning
446 the initial dataset to a training set (including a validation set) and a testing set. In this

447 case the testing set is used solely for establishing the final, unbiased, performance of
448 the classifier (Olden *et al.*, 2008). As this partition reduces the data available for training
449 and validating, partitioning the training dataset further may not be realistic as an
450 inappropriately small training set will create an inappropriate and untrustworthy
451 classifier.

452

453 In order to reduce overfitting during the validating process, cross-validation (CV) can
454 be used instead. In cross-validation the training dataset is partitioned, creating a
455 training set (in the strict sense) and a validation set (Olden *et al.*, 2008). Training
456 commences and is terminated when the performance with respect to the validation
457 set begins to reduce. The validation set is thus used as a dummy 'test' set. After the
458 classifier is trained and validated the two datasets are re-combined and re-partitioned
459 creating a new training and validation dataset. The learning process is repeated again
460 from the start until either a predefined number of data partitions, or all possible data
461 partitions, have been used for training. In biological applications of machine learning,
462 multifold (K-fold) cross-validation is commonly used to help avoid overfitting (Olden *et*
463 *al.*, 2008). During that process the training dataset is partitioned into K equal sets, with K-
464 1 of these recombined to create the training set and the last one used to validate. This
465 process is repeated K times for all possible (or sensible) combinations of training and
466 validation sets. More recently this technique has been slightly modified to include
467 further repetitions; for example, in M repetitions of K-fold cross-validation the process of
468 K-fold cross-validation already described is repeated M times. An example using two
469 repetitions of 5-fold cross-validation is illustrated in Figure 2.

470

471 Throughout this paper, we have explained and illustrated the many strengths of
472 morphometric study including the ability to train and evaluate a system, to conduct
473 power analysis on trial data sets to help decide on appropriate sample sizes and the
474 crucial element of reproducible measurement. Morphometric approaches can offer to

475 build strong and reproducible systems of classification and these can be combined
476 with DNA derived data to give a holistic synthesis that might improve the stability and
477 decrease the subjectivity of plant classification, especially at the species level. In short,
478 when botanists and horticulturalists catch up with other disciplines we expect to see
479 use of morphological data in the construction of more robust botanical classification
480 systems.

481

482 REFERENCES

- 483 **Angermueller C, Pärnamaa T, Parts L, Stegle O. 2016.** Deep learning for computational
484 biology. *Molecular Systems Biology* **12**: 878.
- 485 **Assis LC. 2009.** Coherence, correspondence, and the renaissance of morphology in
486 phylogenetic systematics. *Cladistics* **25**: 528–544.
- 487 **Atay E, Pirlak L, Atay A. 2010.** Determination of fruit growth in some apple varieties.
488 *Journal of Agricultural Sciences* **16**: 1–8.
- 489 **Blanco-Dios JB. 2007.** Estudio morfométrico de una zona híbrida entre *Armeria beirana*
490 y *A. pubigera* (Plumbaginaceae) en el noroeste de la Península Ibérica. *Anales del*
491 *Jardín Botánico de Madrid* **64**: 229–235.
- 492 **Van Bocxlaer B, Schultheiß R. 2010.** Comparison of morphometric techniques for shapes
493 with few homologous landmarks based on machine-learning approaches to biological
494 discrimination. *Paleobiology* **36**: 497–515.
- 495 **Bollard EG. 1970.** The physiology and nutrition of developing fruit. In: Hulme AC, Rhodes
496 MJ, eds. *The biochemistry of fruit and their products: Volume I.*, 387–425.
- 497 **Bonhomme V, Claude J. 2014.** Momocs: Outline analysis using R. *Journal of Statistical*
498 *Software* **56**: 1–24.
- 499 **Christodoulou, M. D., Battey NH, Culham A. 2018.** Can you make morphometrics work
500 when you know the right answer? Pick and mix approaches for apple identification.
- 501 **Clark JY, Corney D, Wilkin P. 2017.** Leaf-based automated species classification using
502 image processing and neural networks. In Lestrel P, ed. *Biological Shape Analysis -*
503 *Proceedings of the 4th International Symposium*: 29-56. World Scientific, Singapore.
- 504 **Compton JA, Hedderson TA. 1997.** A morphometric analysis of the *Cimicifuga foetida* L.
505 complex (Ranunculaceae). *Botanical Journal of the Linnean Society* **123**: 1–23.
- 506 **Corney DPA, Tang HL, Clark JY, Hu Y, Jin J. 2012.** Automating digital leaf measurement:
507 The tooth, the whole tooth, and nothing but the tooth. *PLoS ONE* **7**: 1–10.
- 508 **Da Costa AF, Rodrigues PJFP, Wanderley MDGL. 2009.** Morphometric analysis and
509 taxonomic revision of the *Vriesea paraibica* complex (Bromeliaceae). *Botanical*

510 *Journal of the Linnean Society* **159**: 163–181.

511 **Cuni Sanchez A, De Smedt S, Haq N, Samson R. 2011.** Variation in baobab seedling
 512 morphology and its implications for selecting superior planting material. *Scientia*
 513 *Horticulturae* **130**: 109–117.

514 **Fernández-Mazuecos M, Blanco-Pastor JL, Gómez JM, Vargas P. 2013.** Corolla
 515 morphology influences diversification rates in bifid toadflaxes (*Linaria* sect.
 516 *Versicolores*). *Annals of Botany* **112**: 1705–1722.

517 **Fisher R. 1936.** The use of multiple measurements in taxonomic problems. *Annals of*
 518 *Eugenics* **7**: 179–188.

519 **Fogel GB. 2008.** Computational intelligence approaches for pattern discovery in
 520 biological systems. *Briefings in Bioinformatics* **9**: 307–316.

521 **Gardner ZE, Lueck L, Erhardt EB, Craker LE. 2012.** A morphometric analysis of *Actaea*
 522 *racemosa* L. (Ranunculaceae). *Journal of Medicinally Active Plants* **1**: 47–59.

523 **Giribet G. 2010.** A new dimension in combining data? The use of morphology and
 524 phylogenomic data in metazoan systematics. *Acta Zoologica* **91**: 11–19.

525 **Goodall C. 1991.** Procrustes Methods in the Statistical Analysis of Shape. *Journal of the*
 526 *Royal Statistical Society. Series B (Methodological)* **53**: 285–339.

527 **Guisande C, Manjarrés-Hernández A, Pelayo-Villamil P, Granado-Lorencio C, Riveiro I,**
 528 **Acuña A, Prieto-Piraquive E, Janeiro E, Matías JM, Patti C, Patti B, Mazzola S, Jiménez S,**
 529 **Duque V, Salmerón F. 2010.** Ipez: An expert system for the taxonomic identification of
 530 fishes based on machine learning techniques. *Fisheries Research* **102**: 240–247.

531 **Jenner RA. 2004.** When molecules and morphology clash: Reconciling conflicting
 532 phylogenies of the Metazoa by considering secondary character loss. *Evolution and*
 533 *Development* **6**: 372–378.

534 **Klingenberg CP. 2008.** Novelty and 'homology-free' morphometrics: What's in a name?
 535 *Evolutionary Biology* **35**: 186–190.

536 **Klingenberg CP, Monteiro LR. 2005.** Distances and directions in multidimensional shape
 537 spaces: implications for morphometric applications. *Systematic Biology* **54**: 678–688.

538 **Lagomarsino LP, Forrestel EJ, Muchhala N, Davis CC. 2017.** Repeated evolution of
539 vertebrate pollination syndromes in a recently diverged Andean plant clade. *Evolution*
540 **71**: 1970–1985.

541 **Macleod N. 1999.** Generalizing and extending the eigenshape method of shape space
542 visualization and analysis. *Paleobiology* **25**: 107–138.

543 **McDonald JH. 2014.** *Handbook of Biological Statistics*. Baltimore, Maryland.: Sparky
544 House Publishing.

545 **Nagahama N, Anton AM, Norrmann G a. 2014.** Taxon Delimitation in the *Andropogon*
546 *lateralis* Complex (Poaceae) in Southern South America based on Morphometrical
547 Analyses. *Systematic Botany* **39**: 804–813.

548 **Neves B, Zanella CM, Kessous IM, Uribbe FP, Salgueiro F, Bered F, Antonelli A, Bacon CD,**
549 **Costa AF. 2020.** Drivers of bromeliad leaf and floral bract variation across a latitudinal
550 gradient in the Atlantic Forest. *Journal of Biogeography* **47**: 261–274.

551 **Olden JD, Lawler JJ, Poff NL. 2008.** Machine learning methods without tears: a primer for
552 ecologists. *The Quarterly Review of Biology* **83**: 171–193.

553 **De Oliveira Plotze R, Martinez Bruno O. 2009.** Automatic Leaf Structure Biometry:
554 Computer Vision Techniques and their Applications in Plant Taxonomy. *International*
555 *Journal of Pattern Recognition and Artificial Intelligence* **23**: 247–262.

556 **Perkins SL, Martinsen ES, Falk BG. 2011.** Do molecules matter more than morphology?
557 Promises and pitfalls in parasites. *Parasitology* **138**: 1664–1674.

558 **Richardson AC, Boldingh HL, McAtee PA, Gunaseelan K, Luo Z, Atkinson RG, David KM,**
559 **Burdon JN, Schaffer RJ. 2011.** Fruit development of the diploid kiwifruit, *Actinidia*
560 *chinensis* 'Hort16A'. *BMC Plant Biology* **11**.

561 **Rohlf F, Slice D. 1990.** Extensions of the Procrustes method for the optimal
562 superimposition of landmarks. *Systematic Biology* **39**: 40–59.

563 **Santana FS, Costa AHR, Truzzi FS, Silva FL, Santos SL, Franco TM, Saraiva AM. 2014.** A
564 reference process for automating bee species identification based on wing images
565 and digital image processing. *Ecological Informatics* **24**: 248–260.

566 **Savriama Y, Gómez JM, Perfectti F, Klingenberg CP. 2012.** Geometric morphometrics of
567 corolla shape: Dissecting components of symmetric and asymmetric variation in
568 *Erysimum mediohispanicum* (Brassicaceae). *New Phytologist* **196**: 945–954.

569 **Schneider H, Smith AR, Pryer KM. 2009.** Is morphology really at odds with molecules in
570 estimating fern phylogeny? *Systematic Botany* **34**: 455–475.

571 **Shipunov AB, Bateman RM. 2005.** Geometric morphometrics as a tool for understanding
572 *Dactylorhiza* (Orchidaceae) diversity in European Russia. *Biological Journal of the*
573 *Linnean Society* **85**: 1–12.

574 **da Silva FL, Sella MLG, Franco TM, Costa AHR. 2015.** Evaluating classification and
575 feature selection techniques for honeybee subspecies identification using wing images.
576 *Computers and Electronics in Agriculture* **114**: 68–77.

577 **Stegmann M, Gomez DD. 2002.** A brief introduction to statistical shape analysis.

578 **Tarca AL, Carey VJ, Chen X wen, Romero R, Drăghici S. 2007.** Machine learning and its
579 applications to biology. *PLOS Computational Biology* **3**: 0953–0963.

580 **Vasconcelos TNC, Chartier M, Prenner G, Martins AC, Schönenberger J, Wingler A,**
581 **Lucas E. 2019.** Floral uniformity through evolutionary time in a species-rich tree lineage.
582 *New Phytologist* **221**: 1597–1608.

583 **Velemínská J, Krajíček V, Dupej J, Gómez-Valdés JA, Velemínský P, Šefčáková A,**
584 **Pelikán J, Sánchez-Mejorada G, Brůžek J. 2013.** Technical Note: Geometric
585 morphometrics and sexual dimorphism of the greater sciatic notch in adults from two
586 skeletal collections: The accuracy and reliability of sex classification. *American Journal*
587 *of Physical Anthropology* **152**: 558–565.

588 **Vieira M, Mayo SJ, de Andrade IM. 2014.** Geometric morphometrics of leaves of
589 *Anacardium microcarpum* Ducke and *A. occidentale* L. (Anacardiaceae) from the
590 coastal region of Piauí, Brazil. *Revista Brasileira de Botanica* **37**: 315–327.

591 **Viscosi V, Fortini P, Slice DE, Loy A, Blasi C. 2009.** Geometric morphometric analyses of
592 leaf variation in four oak species of the subgenus *Quercus* (Fagaceae). *Plant*
593 *Biosystems - An International Journal Dealing with all Aspects of Plant Biology* **143**: 575–

594 587.

595 **Volkova PA, Shipunov AB. 2007.** Morphological variation of *Nymphaea*
 596 (*Nymphaeaceae*) in European Russia. *Nordic Journal of Botany* **25**: 329–338.

597 **Walker J. 2000.** Ability of geometric morphometric methods to estimate a known
 598 covariance matrix. *Systematic Biology* **49**: 686–696.

599 **Wilf P, Zhang S, Chikkerur S, Little SA, Wing SL, Serre T. 2016.** Computer vision cracks the
 600 leaf code. *Proceedings of the National Academy of Sciences*: 201524473.

601 **Wolpert DH, Macready WG. 1997.** No free lunch theorems for optimization. *IEEE*
 602 *Transactions on Evolutionary Computation* **1**: 67–82.

603 **Yu X, Liu B, Pei Y, Xu T. 2014.** Evaluation of facial attractiveness for patients with
 604 malocclusion: A machine-learning technique employing Procrustes. *The Angle*
 605 *Orthodontist* **84**: 410–416.

606 **Zelditch ML, Swiderski DL, Sheets HD, Fink WL. 2004.** *Geometric morphometrics for*
 607 *biologists*. Oxford: Elsevier Academic Press.

608 **Zhang L, Ampatzidis Y, Whiting MD. 2015.** Sweet cherry floral organ size varies with
 609 genotype and temperature. *Scientia Horticulturae* **182**: 156–164.

610